Available online at www.sciencedirect.com

ScienceDirect

Procedia - Social and Behavioral Sciences 235 (2016) 537 – 545

Procedia
Social and Behavioral Sciences

12th International Strategic Management Conference, ISMC 2016, 28-30 October 2016, Antalya, Turkey

Model estimation of ARMA using genetic algorithms: A case study of forecasting natural gas consumption

Beyzanur Cayir Ervural[a], Omer Faruk Beyca[b], Selim Zaim[c],b*

[a,b,c] *Istanbul Technical University, Macka, Istanbul, 34357, Turkey*

Abstract

Energy is accepted as a vital strategic issue all over the world due to the important hesitations/concerns about energy reliability, sustainability and affordability. The future of the any country's economy entirely depends on energy because it is the major input and indispensable resource for all sectors. Particularly, natural gas is a common used energy source for electricity generation, heating and cooking. Natural gas dependency on the foreign countries leads to economic damages for developing countries like Turkey, due to the high import costs. In this respect, precise forecasting of natural gas consumption plays crucial role in energy projections and economic progress. Underestimating natural gas demand leads to unsatisfied demand for both industrial and residential needs. In this study, we propose a forecasting method integrating Genetic algorithms (GA) and Autoregressive Moving Average (ARMA) method to take advantages of the unique strength of ARMA and genetic algorithms model. In order to predict natural gas consumption of İstanbul, which is the most important metropolitan city of Turkey, with a lower percentage error and with a greater sensitivity based on penalty function. According to the experimental results, the developed combined approach is more robust and outperforms classical ARMA models in terms of mean absolute percentage error (MAPE) and cost function values.

Peer-review under responsibility of the organizing committee of ISMC 2016.

Keywords: Natural gas consumption; forecasting; genetic algorithms; ARMA

* Corresponding author. Tel.: +90-551-109-4103; fax: +90-212-240-7260.
E-mail address: beyca@itu.edu.tr

Peer-review under responsibility of the organizing committee of ISMC 2016.
doi:10.1016/j.sbspro.2016.11.066

1. Introduction

Natural gas is mostly preferred primary energy resource due to the numerous advantages. Natural gas is an environmental friendly energy source and it can be easily stored and more smoothly transported compared to the other fuels. Additionally, it has a very wide application area such as commercial, industrial, electric power generation and residential applications.

The growing energy demand have resulted in dependency on energy imports, primarily of oil and gas ("From Rep. of Turkey Ministry of Foreign Affairs," 2016). In Turkey, the largest share in the primary energy sources is natural gas with a rate of 35 %, then coal (28, 5 %), hydro (7 %) and other sources (2, 5 %) follow, respectively. According to Ministry of Energy and Natural Resources (MENR), natural gas accounted for 37, 8 % of total electric generation in 2015 (in Fig. 1). In this viewpoint, accuracy of the prediction model of natural gas consumption has emerged as a crucial energy strategy for policy makers and energy authorities in order to minimize economic losses and eliminate undesired conditions (Taşpınar, Çelebi, & Tutkun, 2013).

In the literature, many forecasting studies have been recently applied to formulate natural gas consumption (Ali Azadeh, Saberi, Asadzadeh, Hussain, & Saberi, 2013; Soldo, 2012; Y. Yu, Zheng, & Han, 2014). Forecasting models can be classified into traditional methods, artificial methods and hybrid approaches (F. Yu & Xu, 2014). Time series modelling is one of the most used traditional econometric methods which is widely used in natural gas consumption. Demirel et al. (2012), Erdogdu (2010), Kumar and Jain (2010), Ediger and Akar (2007), Sarak and Satman (2003), Liu and Lin (1991) have applied time series modelling for prediction of natural gas consumption. Artificial methods include Artificial Neural Networks (ANNs) and Genetic Algorithms (GAs). Szoplik (2015), Azadeh et al. (2015), Rodger (2014), Gorucu et al. (2004), Hong et al. (2013), Azadeh et al. (2010), Eynard et al. (2011), Kizilaslan and Karlik (2008) have developed ANN for natural gas prediction, and also Izadyara et al. (2015), Askari et al. (2015), Wang et al. (2012), Thomas et al. (2008) have considered GA approach for natural gas forecasting model, respectively. Hybrid models provide advantages over the single models since combinational methods collect the strong sides of the used methods. Some of the studies include combinational methods as follows: Yu and Xu (2014) used real-coded GA and modified BP neural network, Hong et al. (2013) have conducted GA-support vector machine together, Song and Song (2012) have applied BP neural network with GA on short term gas load prediction. Ong et al. (2005) proposed a GA based model identification to cope with the problem of local optima in ARIMA models and they have concluded better solutions than any ARIMA model.

In this study, we aim to conduct one of the conventional methods is namely ARMA and genetic algorithms, with combining best sides of each method, in order to increase the chance to capture various patterns in the data and improve prediction accuracy (Zhang, 2003). This study will be helpful for energy policy makers with accurate estimation of consumption model by enabling long term planning and investments in the natural gas sector. Different from other works in the literature, this study takes into consideration a fitness function value with reflecting prediction power of the model by considering financial dimension of the model according to over/under estimate.

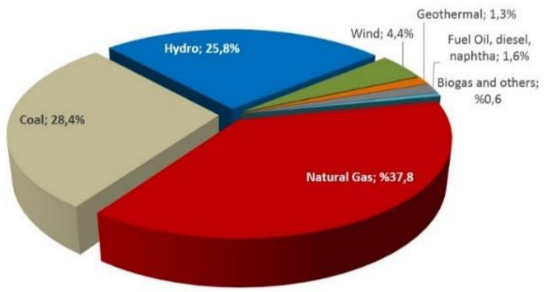

Fig. 1. Electricity generation by type (MENR, 2015)

The rest of this study proceeds as follows. Background information about research methodology is given in Section 2. The proposed method is presented in Section 3. Application of the integrated genetic-ARMA methodology is shown in Section 4. Finally, Section 5 provides a brief conclusion and discussed some future remarks.

2. Background information about research methodology

2.1. Genetic Algorithms

Genetic Algorithms (GA) is an evolutionary optimization approach based on random search algorithms, and developed by Holland in 1975 (Holland, 1992). GA is very popular combinatorial optimization method due to i its robustness for complex and non-linear problems. GA has numerous advantages over other classical optimization methods. In order to obtain better solutions various genetic operators such as selection, mutation and crossover have been implemented to the algorithm. The algorithm can easily converge to good if not the best solution faster than other classical approaches. The basic steps of the GA working principle are as follows: First, the algorithm initialize a population of possible solutions, then applied GA operators which are selection, crossover and mutation operators, respectively. The evaluation function is calculated for each candidate solution. After eliminating bad solution from population, new population is created again by using GA operators and the working mechanism proceed until stopping criterion is satisfied.

2.2. ARMA Models

Box-Jenkins (ARMA) are very popular forecasting methods due to the high level prediction abilities in certain types of data. Box et al. (1994) developed a novel model called ARMA (autoregressive moving average model) in 1994, which is applied to forecast for the stationary time series. ARMA model is a combination of the AR and MA models. In nonstationary series, firstly data should be transformed to stationary form by introducing difference operation. It is the first condition of building an autoregressive integrated moving average model (ARIMA).

In autoregressive moving average model, the future value of a variable is assumed to be a linear function of several past observation and random errors.

$$y_t = \theta_0 + \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots \emptyset_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

where y_t is the actual value and ε_t is random error at the time period t. θ_i (i=1,2,…,p) and $\emptyset_j$ (j=0,1,2,…,q) are model parameters. The order of an ARMA model is usually denoted by the notation ARMA (*p, q*), where *p* is the order of the autoregressive part (AR), *q* is the order of the moving-average process (MA).

3. The proposed method: An integrated genetic-ARMA approach

The principle of genetic algorithms is based on the best individuals' survival. The fitness function improves over generations and the best solution is finally obtained. At first, the initial population is generated randomly. Each element of the population is encoded as a real number in range of [0-1]. After selecting more appropriate individuals according to fitness evaluation for the mating pool, crossover and mutation operators are applied, respectively to produce new offspring. The procedure repeats predefined iteration number times.

3.1. String representation

Each chromosome consists of two parts to represent AR (p) and MA (q), and each dimension equals to p + q length. Each chromosomes is made up of real value coding and the values between 0.0 and 1.0. For instance, an ARMA (2, 4) chromosome can be represented as follows:

AR (2)		*MA (4)*			
0.756	*-0.254*	*0.542*	*0.978*	*-0.076*	*0.854*

3.2. Population Initialization

Initial population is randomly selected. Population size is the number of chromosomes in each generation and it is an important parameter to increase performance of genetic algorithms. There is no standard to specify the size.

3.3. Fitness Function

The aim of this study is to make a prediction according to ARMA model (Eq. 1) with minimum error. Actually, we try to minimize deviations from actual data. The fitness function is created as mean absolute percentage error (MAPE) and calculated as follow function: where D_{actual}, $D_{estimated}$ are actual and estimated natural gas consumption, and n is the number of observations.

$$\min MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(D_{actual(i)} - D_{estimated(i)})}{D_{actual(i)}}\right| \tag{2}$$

Besides this, in this study we try to minimize total cost of the underestimation of natural gas consumption. If the error is less than 0 (error < 0), there will occur a serious unmet demand crisis for the province people for the certain time period. Cost function is calculated as follows

$$cost = \sum_{i} e_i \tag{3}$$

$$e_i = \begin{cases} \left|D_{actual(i)} - D_{estimated}\right|, & D_{actual(i)} \le D_{estimated} \\ p * \left|D_{actual(i)} - D_{estimated}\right|, & D_{actual(i)} > D_{estimated} \end{cases} \tag{4}$$

where p is the penalty value for underestimated demand.

3.4. Selection

Selection is a significant part of the evolutionary algorithm to reach the best chromosomes. The selection operator chooses chromosomes from mating pool according to GA’s working principle, “the fittest individuals have a greater chance of survival than weaker ones”. Roulette wheel is utilized for a probabilistic selection with using following formula:

$$P(x_i) = \frac{f(x_i)}{\sum_{j=1}^{N} f(x_j)} \tag{5}$$

3.5. Crossover

Crossover operator provides new offspring for the next generation with exchanging information between randomly selected two parent chromosomes. Diversification is very important in GA and crossover provides much superiorities to GA in terms of exploration and diversification abilities to achieve global optimum point. In this work, we have used two-point crossover with a crossover probability, P_c.

3.6. Mutation

Mutation operator is utilized to put new genetic information with modifying the genes of a chromosome selected with a mutation probability, Pm. Due to the real value coded, we have added a small number or subtracted a small number from selected value to realize mutation operator according to mutation rate in the study. Mutation is a divergence operation which provides avoiding local optima in the search space.

3.7. Stopping criteria

We used two different stopping criteria. The maximum number of iterations is selected the first termination criteria. Another criteria is decided according to improvement of the fitness function. If there is no improvement on the last improved solution's fitness function after a prescribed number of iterations, then the algorithm is stopped.

3.8. Pseudo code

```
algorithm: GA_ARMA for natural gas forecasting
input: normalized consumption data, GA parameters
output: best solution
begin
    Initialize
            Initialize the GA parameters: population size (N), iteration number, crossover probability (pc),
                    mutation probability (pm), neighborhood rate
            Initialize population (pop) with dimension p + q (pop= 2*rand (k + p, N) -1)
    Evaluate population:  error for each individual (| predicted value –real value |)
    while (not termination condition) do
            Selection: fitness proportionate selection
            Crossover: with probability pc, 2-point crossover operator
            Mutation: with probability pm and neighborhood rate
            if (rand(0,1) < 0.5) increase using neighborhood rate
                    else decrease using neighborhood rate
            Get new population
            Evaluate new population
    end
    output best parameter set for ARMA model
end
```

4. Application of the proposed methodology

The required data is obtained from IGDAS which is a natural gas distributor in Istanbul. The dataset is collected from January 2004 to October 2015, a total of 142 monthly natural gas consumption observations in the residential and commercial areas of Istanbul. The time series plot of the natural gas consumption data is given in Fig. 2.

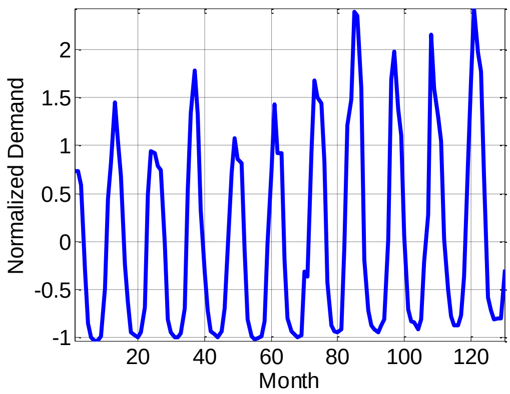

Fig. 2. The time series plot of the natural gas consumption data of Istanbul

The trend effect has been removed by taking the first difference as seen in Fig. 3. After this adjustment on the data, we can apply genetic algorithms for prediction. GA can overcome nonlinear and nonstationary series.

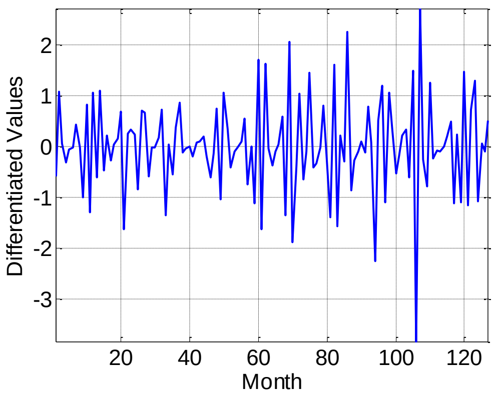

Fig. 3. Time series plot after the first difference of annual natural gas consumption of Istanbul

According to the empirical results, GA-ARMA model outperforms ARMA model due to its dynamic, adaptive and objective structure. Separately ARMA models, could not show an effective performance on prediction of the data due to the traditional structures and subjective judgements on the determination of ARMA parameters. The combined GA-ARMA model fulfill these deficiencies in terms of obtaining better solutions and consistent predictions. As seen in Fig. 4, the developed GA quickly converges to the global/local optimal point and capture the optimum or near optimum solutions easily due to their exploitation and exploration capabilities.

Table 1. Forecasting results

	GA-ARMA	ARMA
MAPE	0.1356	0.1991
Cost Function	12,4660	18,1840

Table 1 gives the forecasting results of the models. According to MAPE and cost function values, GA-ARMA model is much better than ARMA. The proposed approach eliminates lacks of features of the ARMA model individually.

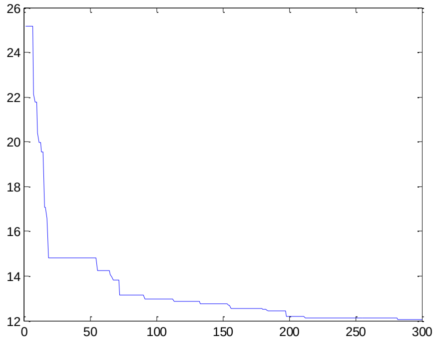

Fig. 4. Convergence plot of GA

Overestimation or underestimation leads to great losses. More precise estimates provide lower costs and better energy resource planning. According to Table 1, GA-ARMA gives a lower penalty cost depending on lower deviation from actual data, it also means accurate prediction. Fig. 5 shows a comparison plot of the utilized forecasting methods according to actual natural gas consumption. The proposed GA-ARMA model provides better and consistent prediction compared to ARMA model.

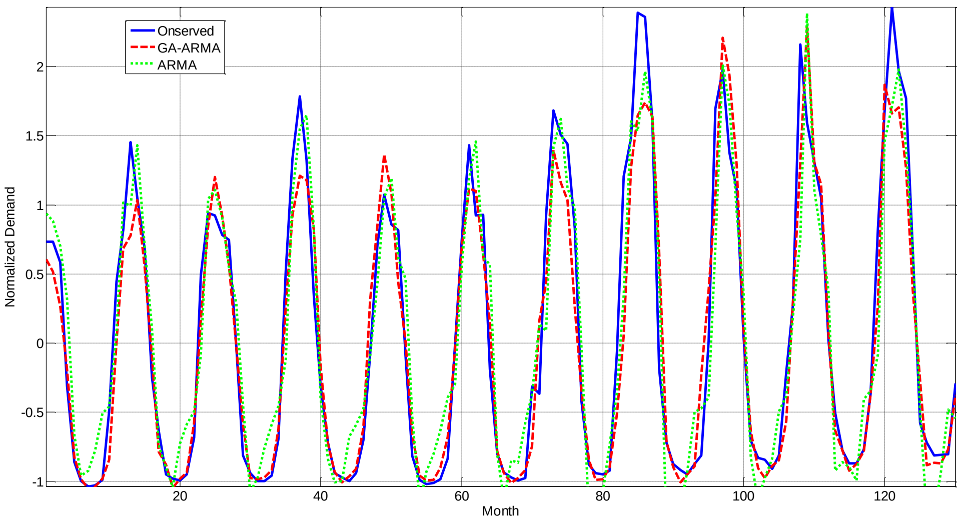

Fig. 5. Comparison of the forecasting models

5. Conclusion

In this paper, we have proposed an integrated forecasting model, which is based on GA and ARMA models, in order to increase accuracy of prediction of natural gas consumption. Overestimate or underestimate is very crucial for energy sector due to the availability of the energy resource. Additionally, all energy resource planning activities and energy investment decisions, which all have high level strategic scope for Turkey, are completely depend on precise forecasting. At this point, conventional methods alone may be insufficient to meet correct forecasting expectations owing to the lack of pattern identification of ARMA and the subjective judgments on the lag determination. In this study, Genetic algorithms provide an objective and effective identification way for parameter estimation of ARMA method. For future directions, other artificial intelligent methods, which are neural networks, support vector machine, can be performed individually or collectively in order to evaluate and compare advantages and deficiencies of the methods to obtain successful and accurate results.

References

Askari, S., Montazerin, N., & Zarandi, M. H. F. (2015). A clustering based forecasting algorithm for multivariable fuzzy time series using linear combinations of independent variables. *Applied Soft Computing*, *35*, 151–160. http://doi.org/10.1016/j.asoc.2015.06.028

Azadeh, A., Asadzadeh, S. M., & Ghanbari, A. (2010). An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: Uncertain and complex environments. *Energy Policy*, *38*(3), 1529–1536. http://doi.org/10.1016/j.enpol.2009.11.036

Azadeh, A., Saberi, M., Asadzadeh, S. M., Hussain, O. K., & Saberi, Z. (2013). A neuro-fuzzy-multivariate algorithm for accurate gas consumption estimation in South America with noisy inputs. *International Journal of Electrical Power & Energy Systems*, *46*, 315–325. http://doi.org/10.1016/j.ijepes.2012.10.013

Azadeh, A., Zarrin, M., Rahdar Beik, H., & Aliheidari Bioki, T. (2015). A neuro-fuzzy algorithm for improved gas consumption forecasting with economic, environmental and IT/IS indicators. *Journal of Petroleum Science and Engineering*, *133*, 716–739. http://doi.org/10.1016/j.petrol.2015.07.002

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time Series Analysis: Box/Time Series Analysis*. Hoboken, NJ: John Wiley & Sons, Inc. Retrieved from http://doi.wiley.com/10.1002/9781118619193

Demirel, Ö. F., Zaim, S., Çalişkan, A., & Özuyar, P. (2012). Forecasting natural gas consumption in Istanbul using neural networks and multivariate time series methods. *Turkish Journal of Electrical Engineering & Computer Sciences*, *20*(5), 695–711.

Ediger, V. Ş., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*, *35*(3), 1701–1708. http://doi.org/10.1016/j.enpol.2006.05.009

Erdogdu, E. (2010). Natural gas demand in Turkey. *Applied Energy*, *87*(1), 211–219. http://doi.org/10.1016/j.apenergy.2009.07.006

Eynard, J., Grieu, S., & Polit, M. (2011). Wavelet-based multi-resolution analysis and artificial neural networks for forecasting temperature and thermal power consumption. *Engineering Applications of Artificial Intelligence*, *24*(3), 501–516. http://doi.org/10.1016/j.engappai.2010.09.003

From Rep. of Turkey Ministry of Foreign Affairs. (2016). Retrieved May 28, 2016, from http://www.mfa.gov.tr/turkeys-energy-strategy.en.mfa

Gorucu, F. B. (2004). Artificial Neural Network Modeling for Forecasting Gas Consumption. *Energy Sources*, *26*(3), 299–307. http://doi.org/10.1080/00908310490256626

Holland, J. H. (1992). *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence* (1st MIT Press ed). Cambridge, Mass: MIT Press.

Hong, W.-C., Dong, Y., Zhang, W. Y., Chen, L.-Y., & K. Panigrahi, B. (2013). Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. *International Journal of Electrical Power & Energy Systems*, *44*(1), 604–614. http://doi.org/10.1016/j.ijepes.2012.08.010

Izadyar, N., Ong, H. C., Shamshirband, S., Ghadamian, H., & Tong, C. W. (2015). Intelligent forecasting of residential heating demand for the District Heating System based on the monthly overall natural gas consumption. *Energy and Buildings*, *104*, 208–214. http://doi.org/10.1016/j.enbuild.2015.07.006

Kizilaslan, R., & Karlik, B. (2008). Comparison neural networks models for short term forecasting of natural gas consumption in Istanbul. In *Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the* (pp. 448–453). http://doi.org/10.1109/ICADIWT.2008.4664390

Kumar, U., & Jain, V. K. (2010). Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy*, *35*(4), 1709–1716. http://doi.org/10.1016/j.energy.2009.12.021

Liu, L.-M., & Lin, M.-W. (1991). Forecasting residential consumption of natural gas using monthly and quarterly time series. *International Journal of Forecasting*, *7*(1), 3–16. http://doi.org/10.1016/0169-2070(91)90028-T

Ong, C.-S., Huang, J.-J., & Tzeng, G.-H. (2005). Model identification of ARIMA family using genetic algorithms. *Applied Mathematics and Computation*, *164*(3), 885–912. http://doi.org/10.1016/j.amc.2004.06.044

Rodger, J. A. (2014). A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings. *Expert Systems with Applications*, *41*(4, Part 2), 1813–1829. http://doi.org/10.1016/j.eswa.2013.08.080

Sarak, H., & Satman, A. (2003). The degree-day method to estimate the residential heating natural gas consumption in Turkey: a case study. *Energy*, *28*(9), 929–939. http://doi.org/10.1016/S0360-5442(03)00035-5

Soldo, B. (2012). Forecasting natural gas consumption. *Applied Energy*, *92*, 26–37. http://doi.org/10.1016/j.apenergy.2011.11.003

Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, *85*, 208–220. http://doi.org/10.1016/j.energy.2015.03.084

Taşpınar, F., Çelebi, N., & Tutkun, N. (2013). Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. *Energy and Buildings*, *56*, 23–31. http://doi.org/10.1016/j.enbuild.2012.10.023

Thomas Ng, S., Skitmore, M., & Wong, K. F. (2008). Using genetic algorithms and linear regression analysis for private housing demand forecast. *Building and Environment*, *43*(6), 1171–1184. http://doi.org/10.1016/j.buildenv.2007.02.017

Wang, J.-J., Wang, J.-Z., Zhang, Z.-G., & Guo, S.-P. (2012). Stock index forecasting based on a hybrid model. *Omega*, *40*(6), 758–766. http://doi.org/10.1016/j.omega.2011.07.008

Yu, F., & Xu, X. (2014). A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Applied Energy*, *134*, 102–113. http://doi.org/10.1016/j.apenergy.2014.07.104

Yu, Y., Zheng, X., & Han, Y. (2014). On the demand for natural gas in urban China. *Energy Policy*, *70*, 57–63. http://doi.org/10.1016/j.enpol.2014.03.032

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. http://doi.org/10.1016/S0925-2312(01)00702-0